

Eksempel på utfyllt

sjekkliste for vurdering av en oversiktsartikkel

Hvordan brukes sjekklisten?

Sjekklisten består av tre deler:

- A: Kan du stole på resultatene?
- B: Hva forteller resultatene?
- C: Kan resultatene være til hjelp i praksis?

I hver del finner du underspørsmål og tips som hjelper deg å svare. For hvert av underspørsmålene skal du krysse av for «ja», «nei» eller «uklart». Valget «uklart» kan også omfatte «delvis».

Om sjekklisten

Sjekklisten er inspirert av: Critical Appraisal Skills Programme (2017). *CASP Checklist: 10 questions to help you make sense of a Systematic Review*. <https://casp-uk.net/casp-tools-checklists/> Hentet: 09.03.2017.

Sjekklisten er laget som et pedagogisk verktøy for å lære kritisk vurdering av vitenskapelige artikler. Hvis du skal skrive en systematisk oversikt eller kritisk vurdere artikler som del av et forskningsprosjekt, anbefaler vi andre typer sjekklister. Se www.helsebiblioteket.no/kunnskapsbasert-praksis/kritisk-vurdering/sjekklister

Har du spørsmål om, eller forslag til forbedring av sjekklisten?
Send e-post til Redaksjonen@kunnskapsbasertpraksis.no.

Kritisk vurdering av:

van Middelkoop, M., Rubinstein, S. M., Kuijpers, T., Verhagen, A. P., Ostelo, R., Koes, B. W. & van Tulder, M. W. (2011). A systematic review on the effectiveness of physical and rehabilitation interventions for chronic non-specific low back pain. *European Spine Journal*, 20(1), 19-39.
<https://doi.org/10.1007/s00586-010-1518-3>

Kritisk vurdering ble gjennomført av: Bente Frisk og Nina Rydland Olsen, Institutt for helse og funksjon, Høgskulen på Vestlandet

Kritisk vurdering er fagfellevurdert av: Lillebeth Larun, Folkehelseinstituttet



kunnskapsbasertpraksis.no



KBP på tvers

Del A: Kan du stole på resultatene?

1. Er formålet med oversikten klart formulert?

Ja – Nei – Uklart

Tips: Formålet bør være klart formulert med hensyn til:

- populasjonen (personene oversikten handler om)
- tiltaket som gis til intervensjonsgruppen
- sammenligningstiltaket som gis til kontrollgruppen
- utfallene (endepunktene/resultatene) som vurderes

Kommentar:

Formålet med oversiktsartikkelen er klart formulert med hensyn til populasjonen, tiltaket, sammenligningstiltaket og utfallene. I denne oversiktsartikkelen oppsummerer forfatterne randomiserte kontrollerte studier som har undersøkt kortsiktig og langsiktig effekt av ulike fysioterapitiltak for personer med kroniske ryggplager med hensyn til utfallene smerte, funksjon, selv-opplevd bedring, retur til arbeid og uønskede hendelser.

Formålet er beskrevet i sammendraget (side 19) og på slutten av bakgrunnen (side 20). Utfall er først beskrevet i metoddelen (side 22). Forfatterne sammenligner alle tiltak mot hverandre, og hvilke tiltak som sammenlignes går først tydelig frem i resultatdelen (side 23 til 34). Det var utfordrende å vurdere om formålet var klart formulert, ettersom utfall og sammenligningstiltak ikke var beskrevet i sammendraget og på slutten av bakgrunnen, slik som er mest vanlig.

Formålet med oversiktsartikkelen var relevant for vårt kliniske spørsmål: Har treningsterapi bedre effekt på retur til arbeid hos voksne pasienter med korsryggsmerter sammenlignet med kognitiv atferdsterapi?

2. Søkte forfatterne etter relevante typer studier?

Ja – Nei – Uklart

Tips: Relevante studier:

- svarer på samme spørsmål som oversikten
- har et egnet studiedesign for å besvare spørsmålet. Eksempler på spørsmålstyper og egnede studiedesign:
 - effekt – randomisert kontrollert studie
 - årsak – kohortstudie
 - diagnostikk – tverrsnittstudie med referansetest
 - prognose – kohortstudie
 - erfaringer – kvalitativ studie
- Er det samsvar mellom kriterier for inklusjon av studiene og formålet til oversikten? Se etter:
 - populasjon
 - tiltak
 - sammenligning
 - utfall

Kommentar:

Det er sannsynlig at forfatterne søkte etter relevante studier som svarte på spørsmålet i denne oversiktsartikkelen, ettersom inklusjonskriteriene (side 20 til 21) samsvarer med formålet til oversiktsartikkelen: hva er effekten av ulike fysioterapi og rehabiliteringstiltak (formulert som «physical and rehabilitation interventions» i artikkelen) for personer med kroniske ryggplager med hensyn til utfallene smerte, funksjon, selv-opplevd bedring, retur til arbeid og uønskede hendelser.

Inklusjonskriterier:

Populasjonen: voksne pasienter med kroniske korsryggsmerter

Tiltak: treningsterapi, ryggskole, TENS*, varme/kulde, laserterapi pasient undervisning, massasje, atferdsterapi, ryggstøtte/korsett, traksjon og tverrfaglig behandling

Utfall: smerteintensitet, funksjon/ryggfunksjon, retur til arbeid og side effekter. Primærutfall var retur til arbeid og funksjon.

Design: RCT.

Forfatterne oppgir (side 20) at de søkte etter randomiserte kontrollerte studier (RCT) studier, som er et egnet studiedesign for å svare på effektspørsmål.

*TENS=transcutaneous electrical nerve stimulation.

3. Er det sannsynlig at alle viktige og relevante studier ble funnet?

Ja – Nei – Uklart

Tips: Se etter:

- hvilke databaser og eventuelle søkemotorer forfatterne har søkt i, og hvordan de søkte i dem (søkestrategi)
- om de søkte etter ikke-publiserte studier (i for eksempel Trials.gov)
- om det var noen vesentlige begrensninger med hensyn til språk eller tid (hvis kun studier på engelsk er tatt med, kan det gi skjevhet (bias))

Kommentar:

Det er sannsynlig at alle viktige og relevante studier ble funnet, ettersom forfatterne både screenet tidligere Cochrane reviews for studier som har undersøkt effekten av de 11 aktuelle tiltak i oversiktsartikkelen og gjennomførte søk i databasene MEDLINE, EMBASE, CINAHL, CENTRAL og PEDro. Søkene ble sist oppdatert 22. desember 2008. Forfatterne brukte en tidligere søkestrategi utviklet av the Cochrane Back Review Group. Denne informasjon om søkemotoder er oppgitt i avsnittet «Search methods for identification of studies» (side 21). Ytterlige detaljer om søkestrategien er ikke oppgitt, for eksempel kunne flere detaljer om søkestrategien vært spesifisert i et vedlegg.

Referanselister i relevante publikasjoner ble screenet for aktuelle enkeltstudier, og eksperter ble kontaktet for å identifisere enda flere mulige studier. To forfattere gjennomførte selvstendige søk.

Forfatterne oppgir ikke om de har søkt etter ikke-publiserte studier for eksempel i ClinicalTrials.gov, og dette kan ha resultert i seleksjonsskjevhet. Søkene var begrenset til engelsk, nederlandsk og tysk, og dette kan også ha ført til seleksjonsskjevhet. Disse begrensningene gjør at det er noe usikkert knyttet til vurdering om alle viktige og relevante studier ble funnet.

Av de 83 inkluderte studiene i denne oversiktsartikkelen, var det ingen som besvarte alle deler av vårt kliniske spørsmål. Etter å ha vurdert treffene fra litteratursøket som ble gjennomført for vårt kliniske spørsmål, var det likevel denne oversiktsartikkelen som best svarte på spørsmålet.

4. Ble kvaliteten på de inkluderte studiene tilstrekkelig vurdert?

Ja – Nei – Uklart

Tips: Forfatterne må vurdere den metodiske kvaliteten på enkeltstudiene på en fornuftig måte med gode verktøy/sjekklister. Systematiske skjevheter (bias) i hvordan studiene er utført kan påvirke resultatene i studiene, og dermed også resultatene og konklusjonene i oversikten.

- Er det oppgitt hvilke verktøy/sjekklistet som ble brukt i vurderingen?
- Er verktøyet/sjekklisten som ble brukt fornuftig?
- Ble vurderingen gjort av minst to personer uavhengig av hverandre?

Kommentar:

Metodisk kvalitet på de inkluderte studiene ble tilstrekkelig vurdert. For å vurdere metodisk kvalitet brukte forfatterne et verktøy utarbeidet og anbefalt av the Cochrane Back Review Group. Vi må anta at sjekklistet fra Cochrane grupper er anerkjente. Sjekklisten som ble brukt i dette tilfellet er kjent og publisert. Det er 11 spørsmål i sjekklisten og disse beskrives i oversiktsartikkelen i Table 1, side 24 til 26. Table 1 beskriver vurderingen av den metodiske kvaliteten på de 83 inkluderte studiene. Vurdering av metodisk kvalitet blir presentert på en oversiktlig måte i resultatdelen.

To forfattere gjennomførte vurdering av metodisk kvalitet, uavhengig av hverandre. Eventuelle uenigheter ble diskutert på konsensusmøter.

5. Hvis resultater fra de inkluderte studiene er slått sammen statistisk i en metaanalyse, var dette fornuftig og forsvarlig?

Ja – Nei – Uklart

Tips:

- Er det klart presisert hvilke sammenligninger som ble gjort?
- Kommer resultatene (effekttestimatene med konfidensintervaller) fra enkeltstudiene klart frem?
- Er enkeltstudiene like nok til å slås sammen? (Har forfatterne for eksempel beskrevet metoder for å teste heterogenitet?)
- Er analysemetodene som er brukt forklart? (For eksempel random eller fixed effect)
- Hvis det ikke er gjort en metaanalyse, er resultatene fra de inkluderte studiene narrativt oppsummert? Var det i så fall fornuftig og forsvarlig?

Kommentar:

Resultatene fra de inkluderte studiene er slått sammen statistisk i en metaanalyse. Det er klart presisert hvilke sammenligninger som ble gjort og alle sammenligningene er beskrevet i resultatdelen og illustrert ved hjelp av forestplots i et vedlegg (Supplementary material 2).

I resultatdelen og i forestplots er det oppgitt effekttestimat med konfidensintervaller for hver enkelt studie. Gjennomsnittlig forskjell (Mean difference) er oppgitt som effekttestimat.

For å undersøke om enkeltstudiene er like nok til å slås sammen har forfatterne gjennomført statistiske tester for heterogenitet ved bruk av Q-test/Chi-square og I^2 . I^2 indikerer hvor mange prosent variasjon mellom studiene som utgjør totalvariasjonen, som er summen av variasjon innenfor studiene og variasjonen mellom studiene. Dersom prosent variasjon er $> 50\%$ regnes det som høy grad av heterogenitet.

Resultatene i oversiktsartikkelen viste at det var høy grad av heterogenitet for flere sammenligner som ble gjort, og disse resultatene må derfor tolkes med forsiktighet. Dette gjaldt for eksempel 1) effekten av treningsterapi versus atferdsterapi i forhold til langtidseffekten på funksjon ($I^2=74\%$), 2) effekten av treningsterapi versus TENS/laserterapi/ultraterapi i forhold til smerte ($I^2=77\%$) og funksjon ($I^2=65\%$), 3) effekten av treningsterapi versus manuell terapi/manipulering i forhold til funksjon ($I^2=67\%$), og smerte ($I^2=85\%$).

Det er derfor uklart om enkeltstudiene er like nok til å slås sammen og det er uklart om det var fornuftig og forsvarlig å slå sammen studiene i en statistisk metaanalyse.

Analysemetodene som er brukt er forklart. Forfatterne har brukt random effect model og det ble gjennomført funnel plots for å undersøke for publikasjonsbias.

Basert på svarene dine på punkt 1–5 over, mener du at resultatene fra denne oversikten er til å stole på?

Ja – Nei – Uklart

Del B: Hva forteller resultatene?

6. Hva er resultatene?

Tips: Vurder

- hvor godt du forstår hovedkonklusjonen i oversikten
- hvordan resultatene er fremstilt, se etter:
 - gjennomsnittsforskjell (mean difference)
 - standardisert gjennomsnittlig forskjell (standardised mean difference)
 - numbers needed to treat (NNT)
 - numbers needed to harm (NNH)
 - odds ratio (OR)
 - relativ risiko (RR)
 - relativ risikoreduksjon (RRR)
 - absolutt risikoreduksjon (ARR)

Kommentar:

I denne oversiktsartikkelen ble det inkludert 83 randomiserte kontrollerte studier. Forfatterne oppsummerer resultatene for 11 ulike tiltak (tiltak beskrevet i spørsmål 2). Forfatterne oppsummerer på slutten av sammendraget og begynnelsen av diskusjonen. Under presenteres de viktigste funnene som forfatterne konkluderer med:

Treningsterapi: Forfatterne konkluderer med at sammenlignet med ingen behandling/venteliste hadde treningsterapi en kortsiktig effekt på smertereduksjon og grad av uførhet, og en langvarig effekt på funksjon.

Atferdsterapi: Alle former for atferdsterapi sammenlignet med ingen behandling/venteliste var mer effektiv for å redusere smerteintensitet, men usikker om effekt på funksjon.

Tverrfaglig behandling: hadde en kortsiktig effekt på smertereduksjon sammenlignet med ingen behandling/venteliste.

TENS: Ingen forskjeller i effekt når TENS ble sammenlignet med narre-TENS (sham TENS) eller sammenlignet med aktiv behandling.

Forfatterne beskriver at det er stor heterogenitet mellom studiepopulasjonene, tiltakene og sammenligningsgruppene, og det er derfor utilstrekkelige data til å trekke klare konklusjoner på effekten av ryggskoler, laserterapi, pasientopplæring, massasje, traksjon varme-/kuldebehandling og korsett.

Resultatene er fremstilt med vektet gjennomsnittlig forskjell (weighted mean difference (WMD)) og konfidensintervall (CI). I metaanalyser vil det for kontinuerlige variabler være nødvendig å kombinere målinger hvor gjennomsnitt, standardavvik (SD) og antall deltakere (sample size) i hver gruppe er kjent. Hver enkelt studie blir vektet i forhold til hvor mye innflytelse den har på det totale resultat i metaanalysen (WMD). WMD er en metode som antar at alle studier har målt utfallet på den samme skala, men det var ikke tilfellet med studiene som ble slått sammen her. Alle skalaer ble derfor

omgjort til en 100 poengs skala. Resultatene refererer derfor ikke til en bestemt skala, som for eksempel VAS-skala.

Number needed to treat (NNT), number needed to harm (NNH), odds ratio (OR), relative risk (RR), relative risk reduction (RRR) og absolute risikoreduksjon (ARR) er ikke oppgitt. I supplementary material er forest plot vist for de ulike analysene.

7. Hvor presise er resultatene?

Tips: Se på

- konfidensintervall (KI/CI)
- interkvartilbredde (interquartile range (IQR))

Kommentar:

Resultatene er som nevnt over oppgitt som WMD og CI. For eksempel, trening sammenlignet med standard behandling/råd om å være aktiv viste en statistisk signifikant effekt av trening med reduksjon av smerte rapportert som WMD=-9,23 og 95 % CI -16,02; -2,43 og uførhet WMD=-12,35 og 95 % CI -23,00; -1,69. En endring på -9,23 og -12,35 betyr at smerten ble redusert med 9,23 % og uførhet med 12,35 %. Disse konfidensintervallene krysser ikke null og derfor er det sannsynlig at resultatet er statistisk signifikant (gjelder kontinuerlige variabler), men sett i lys av at dette er en 100 poengs skala er konfidensintervallene brede, noe som indikerer at resultatene er usikre og at vi ikke kan være sikker på at det er en forskjell mellom tiltakene.

Del C: Kan resultatene være til hjelp i praksis?

8. Kan resultatene overføres til praksis?

Ja – Nei – Uklart

Tips: Vurder om, og i hvilken grad

- populasjonen (personene) i studiene som inngår i oversikten er representative for de du møter i praksis
- din praksis skiller seg fra den som omtales i oversikten

Kommentar:

Resultatene kan overføres til praksis, fordi pasientene som inngår i oversiktsartikkelen er representative for personer med korsryggplager, og populasjonen er lik den i vårt kliniske spørsmål. Men det er uklart hvor nyttig resultatene er for pasienter.

Praksis i Norge for behandling av denne pasientgruppen inkluderer både trening og tverrfaglig tilnærming, i tråd med forfatterens konklusjoner om hva som er mest lovende tiltak. Resultatene er dermed relevante, og anvendelse av resultatene vil ikke innebære store omlegginger av praksis. Men det er ingen av de inkluderte tiltakene som både vil redusere smerte, øke funksjonsnivået, redusere uførhet og føre til rask retur av arbeid for denne pasientgruppen.

I tillegg er det uklart om resultatene fører til en klinisk meningsfull endring for pasienter. For å vurdere om det er en klinisk meningsfull endring trenger vi informasjon om hva som er minste klinisk viktig endring, på engelsk, minimal clinically important difference (MCID) for hvert måleinstrument. Forfatterne har ikke oppgitt noe om MCID i metodedelen, men i diskusjonen (side 34, andre kolonne) står det at ingen av de signifikante resultatene oppnådde en endring større enn 10 %, mens i de fleste studier er en endring mellom 15 til 20 % definert som klinisk meningsfull.

Som nevnt over, er det også stor heterogenitet mellom flere av de inkluderte studiene og resultatene må derfor tolkes med forsiktighet.

For å vurdere om resultatene kan overføres til praksis trengs også informasjon om tiltakene i studiene som er inkludert i oversiktsartikkelen.

For å finne ut om tiltakene i de inkluderte studiene tilsvarer det som gjøres i din praksis, må det sjekkes om tiltakene er beskrevet i de inkluderte studiene. Om de er beskrevet (ofte er ikke det gjort), er det viktig at å ha et kritisk blikk på hva tiltakene inneholder. Er øvelser, dosering og progresjon oppgitt slik at treningen kan gjennomføres tilsvarende det som er gjort i de ulike studiene? Er tiltakene hensiktsmessige med hensyn til pasientens funksjonsproblem?

Det kan stilles spørsmål til om det i det hele tatt vil være forventet å finne en effekt når trening sammenlignes med for eksempel andre former for trening (side 28). Ville det ikke vært naturlig å tro at ulike treningstiltak kan gi effekt?

9. Ble alle viktige utfallsmål vurdert?

Ja – Nei – Uklart

Tips: Vurder om, og i hvilken grad det finnes ytterligere informasjon som du ville hatt med i oversikten.

Kommentar:

Smerteintensitet, funksjon/ryggfunksjon, retur til arbeid og side effekter var utfallsmål som ble vurdert og alle disse er viktige utfallsmål når effekten av tiltak for korsryggplager undersøkes. Primærutfall var retur til arbeid og funksjon.

Det kliniske spørsmålet som vi ønsket svar på var: Har treningsterapi bedre effekt på retur til arbeid hos voksne pasienter med korsryggsmarter sammenlignet med kognitiv atferdsterapi?

Retur til arbeid var utfallsmål i vårt kliniske spørsmål og det er også beskrevet som et av utfallsmålene i inklusjonskriteriene i oversiktsartikkelen. I de fleste identifiserte og inkluderte studiene var imidlertid smerte og uførhet de mest vanligste utfallsmålene som ble undersøkt.

Inkluderte studier med retur til arbeid eller sykefravær som utfallsmål:

To studier rapporterte på retur til arbeid og sykefravær (side 32, første kolonne). I den ene studien viste resultatene at 48 % i intervensjonsgruppen (fikk psykologisk behandling i tillegg til standard rehabiliteringsbehandling) hadde returnert til arbeid etter 6 måneder, sammenlignet med 67 % i kontrollgruppen (fikk standard rehabiliteringsbehandling). Forskjellen var ikke statistisk signifikant. I den andre studien var kostnadene i for bindelse med sykefravær lavere i intervensjonsgruppen sammenlignet med den gruppen som fikk standard behandling.

Et annet spørsmål som ble undersøkt var tverrfaglig behandling sammenlignet med ingen behandling/venteliste kontroller. Her viste resultatene fra en studie (side 33) at det var statistisk

signifikant forskjell på sykefraværsdager ved fire måneder, hvor intervensjonsgruppen var sykemeldt 10 dager (median), mens tilsvarende tall for kontrollgruppen var 122 dager.

Et tredje spørsmål var tverrfaglig behandling versus andre typer aktiv behandling (trening, fysioterapi og treningsterapi) (side 34).

En studie hadde smerte, uførhet og retur til arbeid etter fem år som utfallsmål, og de fant ingen forskjell på smerteintensitet, men intervensjonsgruppen hadde mindre uførhet sammenlignet med treningsgruppen. Det er ikke skrevet spesifikt om retur til arbeid.

10. Veier fordelene opp for ulemper og kostnader?

Ja – Nei – Uklart

Tips: Vurder om, og i hvilken grad

- nytten av tiltaket er verd kostnader og eventuelle bivirkninger
- det er praktiske eller organisatoriske ulemper for deg eller pasienten

Kommentar:

Forfatterne konkluderte med at det er lav tillit til dokumentasjonen for effekten av tiltakene i denne oversiktsartikkelen, men tverrfaglig behandling, atferdsterapi og trening foretrekkes som konservativ behandling ved kroniske korsryggsmerter.

Dersom disse tiltakene medfører redusert sykefravær, kan de være kostnadsbesparende, men det er få studier som hadde sett på dette utfallsmålet, og dermed vanskelig å si om tiltakene er verdt kostnadene. Samtidig har de foreslåtte tiltakene ikke rapporterte uønskede hendelser, og dermed kan det ikke gjøre skade å prøve ut disse tiltakene. Tverrfaglig behandling kan være utfordrende å få til i små kommuner. På bakgrunn av dette har vi krysset av for både ja og uklart.